

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-058908

(43)Date of publication of application : 28.02.2003

(51)Int.Cl.

G06T 15/70

G06T 11/60

G10L 13/00

G10L 21/06

(21)Application number : 2001-244176

(71)Applicant : MINOLTA CO LTD

(22)Date of filing : 10.08.2001

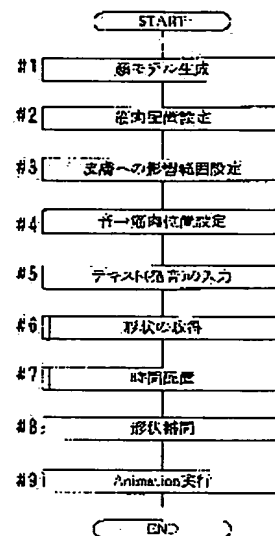
(72)Inventor : TOYAMA OSAMU

(54) METHOD AND DEVICE FOR CONTROLLING FACE IMAGE, COMPUTER PROGRAM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To reduce quantity of data regarding the shape and to realistically realize lip sync animation by control of a face image in comparison with the conventional manner.

SOLUTION: Pieces of first shape data regarding the shapes of a mouth when vowels are uttered are stored by every kind of vowels, kinds of consonants having common points in the shapes of the mouth when the consonants are uttered are sorted into the same groups, second shape data regarding the shapes of the mouth when the consonants sorted into the groups are uttered, are stored (#4) by every group, sounds of a language are sectioned (#5) by every vowel or consonant and motions of the face image are controlled (#8) on the basis of pieces of the first shape data corresponding to the vowels or pieces of the second shape data corresponding to the groups into which the consonants are sorted by every sectioned vowel and consonant.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-58908

(P2003-58908A)

(43) 公開日 平成15年2月28日 (2003.2.28)

(51) Int.Cl. ⁷	識別記号	F I	テマコード [*] (参考)
G 0 6 T 15/70		G 0 6 T 15/70	B 5 B 0 5 0
11/60	2 0 0	11/60	2 0 0 5 D 0 4 5
G 1 0 L 13/00		G 1 0 L 3/00	S
21/06			

審査請求 未請求 請求項の数9 O L (全 13 頁)

(21) 出願番号 特願2001-244176(P2001-244176)

(22) 出願日 平成13年8月10日 (2001.8.10)

(71) 出願人 000006079

ミノルタ株式会社

大阪府大阪市中央区安土町二丁目3番13号

大阪国際ビル

(72) 発明者 遠山 修

大阪府大阪市中央区安土町二丁目3番13号

大阪国際ビル ミノルタ株式会社内

(74) 代理人 100086933

弁理士 久保 幸雄

Fターム(参考) 5B050 AA08 BA08 BA12 EA13 EA24

FA02 FA10

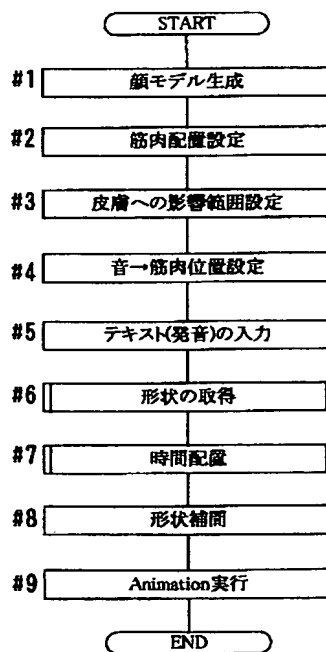
5D045 AB01 AB11

(54) 【発明の名称】 顔画像制御方法および装置、コンピュータプログラム、および記録媒体

(57) 【要約】

【課題】 従来よりも形状に関するデータの量を少なくし、かつ、顔画像の制御によるリップシンクアニメーションをリアルに実現する。

【解決手段】 母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、発音するときの口の形状に共通点がある子音の種類同士を同じグループに分類し、前記グループごとに当該グループに分類された子音を発するときの口の形状に関する第二の形状データを記憶し (#4)、言葉の音を母音または子音ごとに区切り (#5)、区切られた母音または子音ごとに、母音に対応する第一の形状データまたは子音が分類されたグループに対応する第二の形状データに基づいて顔画像の動作の制御を行う (#8)。



【特許請求の範囲】

【請求項 1】言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御方法であって、

母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、

発音するときの口の形状に共通点がある子音の種類同士を同じグループに分類し、前記グループごとに当該グループに分類された子音を発するときの口の形状に関する第二の形状データを記憶し、

前記言葉の音を母音または子音ごとに区切り、

区切られた前記母音または前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類されたグループに対応する第二の形状データに基づいて前記顔画像の動作の制御を行う、

ことを特徴とする顔画像制御方法。

【請求項 2】言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御方法であって、

母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、

発音するときの口の形状が、唇を合わせたようになる特徴を有する第一のグループ、やや開いたようになる特徴を有する第二のグループ、およびその直前の音韻の口の形状のままとなる特徴を有する第三のグループ、の各グループごとに、それぞれの特徴の口の形状に関する第二の形状データを記憶し、

各子音の種類を、当該子音を発するときの口の形状が前記各特徴のうち最も近い特徴を有する前記グループに分類し、

前記言葉の音を母音または子音ごとに区切り、

区切られた前記母音または前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類される前記グループに対応する第二の形状データに基づいて前記顔画像の動作の制御を行う、

ことを特徴とする顔画像制御方法。

【請求項 3】前記言葉の音に所定の連続する 2 つの母音または子音が含まれる場合は、当該連続する 2 つの母音または子音のうち後の母音または子音を所定の母音または子音に置換え、

置き換えられた音について前記制御を行う、

請求項 1 記載の顔画像制御方法。

【請求項 4】母音のときの口の形状を保つ時間よりも子音のときの口の形状を保つ時間のほうが短くなるように前記制御を行う請求項 1 または請求項 3 記載の顔画像制御方法。

【請求項 5】前記言葉の音に所定の連続する 2 つの母音または子音が含まれる場合は、当該連続する 2 つの母音または子音のうちいずれか 1 つまたは両方について、口の形状を保つ既定の時間を変更して前記制御を行う請求項 1、請求項 3、または請求項 4 のいずれかに記載の顔画像制御方法。

【請求項 6】言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御装置であって、

母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データ、および発音するときの口の形状に共通点がある子音の種類が分類されたグループごとに当該グループの子音を発するときの口の形状に関する第二の形状データを記憶する形状データ記憶手段と、前記言葉の音を母音ごとおよび子音ごとに区切り、区切られた前記母音および前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類された前記グループに対応する前記第二の形状データに基づいて前記顔画像の動作の制御を行う動作制御手段と、を有することを特徴とする顔画像制御装置。

【請求項 7】言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御装置であって、

母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、発音するときの口の形状が、唇を合わせたようになる特徴を有する第一のグループ、やや開いたようになる特徴を有する第二のグループ、およびその直前の音韻の口の形状のままとなる特徴を有する第三のグループ、の各グループごとに、それぞれの特徴の口の形状に関する第二の形状データを記憶し、各子音の種類を、当該子音を発するときの口の形状が前記各特徴のうち最も近い特徴を有する前記グループに分類して記憶する形状データ記憶手段と、

前記言葉の音を母音または子音ごとに区切り、区切られた前記母音および前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類された前記グループに対応する前記第二の形状データに基づいて前記顔画像の動作の制御を行う動作制御手段と、を有することを特徴とする顔画像制御装置。

【請求項 8】母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データ、および発音するときの口の形状に共通点がある子音の種類が分類されたグループごとに当該グループの子音を発するときの口の形状に関する第二の形状データを記憶する形状データ記憶手段を有し、言葉に合わせて口が動くように顔画像の動作を制御するコンピュータに用いられるコンピュータプログラムであって、

前記言葉の音を母音ごとおよび子音ごとに区切るステップと、

区切られた前記母音および前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類された前記グループに対応する前記第二の形状データに基づいて前記顔画像の動作の制御を行うステップと、をコンピュータに実行させるためのコンピュータプログラム。

【請求項 9】請求項 8 に記載のコンピュータプログラムが記録されたコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、言葉に合わせて口が動く顔画像いわゆるリップシンクアニメーションの制御方法および制御装置に関する。

【0002】

【従来の技術】従来より、言葉に合わせて顔画像の口を動かし、その顔画像がその言葉を喋っているかのように見せる技術が提案されている。

【0003】例えば、言葉の長さに応じて単に口の開閉を繰り返し行うように顔画像を制御する方法がある。係る方法は、顔画像の制御が簡単である。しかし、すべての音について口の形状が同じになるので、リアル性に乏しい。

【0004】そこで、言葉に合わせて口の形状を変化させるリップシンクアニメーション (Lip Synchronization Animation) の技術が提案されている。例えば、特表 2000-507377号に開示されている発明によると、その言葉 (単語) に含まれる隣り合う 2つの発音表記の組み合わせ、特に、母音と子音との組み合わせに応じて口の形状を制御する。

【0005】また、よりリアルなアニメーションを実現するためには、実際に撮影された映像から Inverse Kinematics などの手法により形状のデータを取得する方法が用いられる。

【0006】

【発明が解決しようとする課題】しかし、特表 2000-507377号の発明によると、単に口の開閉を繰り返し行う方法よりもリアル性は向上するが、発音表記の組み合わせの数が多いので、形状に関するデータの量が多くなる。Inverse Kinematics の手法を用いる場合も、映像から得られる形状が非常に多いので、形状に関するデータの量が多くなる。

【0007】データ量が多くなると、リップシンクアニメーションを実現するために、より高速なネットワークおよびより高性能なコンピュータなどが必要となる。したがって、リップシンクアニメーションの適用範囲を広げるためには、リップシンクアニメーションに必要なデータ量を少なくしなければならない。

【0008】本発明は、このような問題点に鑑み、従来よりも形状に関するデータの量を少なくし、かつ、顔画像の制御によるリップシンクアニメーションをリアルに実現することができる顔画像制御方法および顔画像制御装置を提供することを目的とする。

【0009】

【課題を解決するための手段】本発明に係る顔画像制御方法は、言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御方法であって、母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、発音するときの口の形状に共通点がある子音の種類同士を同じグループに分類し、前記グループご

とに当該グループに分類された子音を発するときの口の形状に関する第二の形状データを記憶し、前記言葉の音を母音または子音ごとに区切り、区切られた前記母音または前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類されたグループに対応する第二の形状データに基づいて前記顔画像の動作の制御を行う。

【0010】好ましくは、言葉に合わせて口が動くように顔画像の動作を制御する顔画像制御方法であって、母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、発音するときの口の形状が、唇を合わせたようになる特徴を有する第一のグループ、やや開いたようになる特徴を有する第二のグループ、およびその直前の音韻の口の形状のままとなる特徴を有する第三のグループ、の各グループごとに、それぞれの特徴の口の形状に関する第二の形状データを記憶し、各子音の種類を、当該子音を発するときの口の形状が前記各特徴のうち最も近い特徴を有する前記グループに分類し、前記言葉の音を母音または子音ごとに区切り、区切られた前記母音または前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類される前記グループに対応する第二の形状データに基づいて前記顔画像の動作の制御を行う。

【0011】または、前記言葉の音に所定の連続する 2つの母音または子音が含まれる場合は、当該連続する 2つの母音または子音のうち後の母音または子音を所定の母音または子音に置換え、置き換えられた音について前記制御を行う。

【0012】または、母音のときの口の形状を保つ時間よりも子音のときの口の形状を保つ時間のほうが短くなるように前記制御を行ってもよい。または、前記言葉の音に所定の連続する 2つの母音または子音が含まれる場合は、当該連続する 2つの母音または子音のうちいずれか 1つまたは両方について、口の形状を保つ既定の時間を変更して前記制御を行ってもよい。

【0013】本発明に係る顔画像制御装置は、母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データを記憶し、発音するときの口の形状が、唇を合わせたようになる特徴を有する第一のグループ、やや開いたようになる特徴を有する第二のグループ、およびその直前の音韻の口の形状のままとなる特徴を有する第三のグループ、の各グループごとに、それぞれの特徴の口の形状に関する第二の形状データを記憶し、各子音の種類を、当該子音を発するときの口の形状が前記各特徴のうち最も近い特徴を有する前記グループに分類して記憶する形状データ記憶手段と、前記言葉の音を母音または子音ごとに区切り、区切られた前記母音および前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類された前記グループに対応する前記第二の形状データに基づいて前記顔画像の動作の制御

を行う動作制御手段と、を有する。

【0014】本発明に係るコンピュータプログラムは、母音の種類ごとに当該母音を発するときの口の形状に関する第一の形状データ、および発音するときの口の形状に共通点がある子音の種類が分類されたグループごとに当該グループの子音を発するときの口の形状に関する第二の形状データを記憶する形状データ記憶手段を有し、言葉に合わせて口が動くように顔画像の動作を制御するコンピュータに用いられるコンピュータプログラムであって、前記言葉の音を母音ごとおよび子音ごとに区切るステップと、区切られた前記母音および前記子音ごとに、当該母音に対応する前記第一の形状データまたは当該子音が分類された前記グループに対応する前記第二の形状データに基づいて前記顔画像の動作の制御を行うステップと、をコンピュータに実行させる。本発明に係る記録媒体は、前記コンピュータプログラムが記録されている。

【0015】

【発明の実施の形態】図1は本発明に係る顔画像制御装置1の構成を説明する図、図2は磁気記憶装置12に記憶されるプログラムおよびデータを示す図、図3は顔画像制御装置1の機能的構成を説明する図である。

【0016】図1に示すように、顔画像制御装置1は、処理装置10、ディスプレイ装置11、磁気記憶装置12、キーボード13、マウス14、マイク15、およびスピーカ16などによって構成される。

【0017】処理装置10は、CPU10a、RAM10b、ROM10c、各種の入出力ポート10d、および各種のコントローラ10eなどによって構成される。磁気記憶装置12には、図2に示すように、オペレーティングシステム(OS)12a、顔画像制御プログラム12b、およびモデリングプログラム12cなどのプログラム、および後に説明する種々の処理に用いられるデータなどが記憶されている。

【0018】磁気記憶装置12に記憶されているプログラムおよびデータは、必要に応じてRAM10bにロードされる。ロードされたプログラムは、CPU10aによって実行される。ネットワーク6Nを介して顔画像制御装置1を他のコンピュータに接続し、プログラムまたはデータをダウンロードしてもよい。または、フロッピーディスク19a、CD-ROM19b、または光磁気ディスク(MO)19cなどの各種リムーバブルディスク(記録媒体)からプログラムまたはデータをロードしてもよい。

【0019】ディスプレイ装置11には、処理装置10による処理結果が表示される。例えば、入力された言葉に合わせて口が動くように人物の顔画像HFを制御し、係る制御の結果をリップシンクアニメーションとしてディスプレイ装置11の表示画面HGに表示する。スピーカ16は、顔画像HFの動作(リップシンクアニメーション)

ン)に合わせてその言葉を音声として出力する。これにより、顔画像HFがその言葉を喋っているかのようにユーザに認識させることができる。キーボード13またはマイク15は、顔画像HFに喋らせる言葉を入力することなどに用いられる。

【0020】顔画像制御装置1として、例えば、ワークステーションまたはパーソナルコンピュータなどが用いられる。このような構成によって、顔画像制御装置1には、図3に示すように、顔モデルデータ生成部101、顔モデルデータ記憶部102、テキストデータ取得部103、音声データ取得部104、音声テキスト変換部105、顔画像制御部106、形状変化データ記憶部107、形状グループ記憶部108、および音声出力部109などが設けられる。

【0021】人物の顔画像HFは、その人物の3次元形状モデルを所定の方向から2次元上に投影することによって得られる。つまり、顔画像HFの動作は、3次元形状モデルを言葉に合わせて変形することによって制御される。

〔3次元形状モデルの準備〕図4は3次元形状モデルの生成の処理の流れを説明するフローチャート、図5は標準モデルDSの例を示す図、図6は変形処理の流れを説明するフローチャート、図7は標準モデルDSの面Sと3次元計測データの点Pとを模式的に示す図、図8は標準モデルDSの異常変形を防ぐための仮想バネを説明するための図である。

【0022】図3の顔モデルデータ生成部101は、上に述べた顔画像HFの基となる3次元形状モデルを生成する。3次元形状モデルの生成は、図4に示すフローチャートのような手順で行われる。

【0023】まず、図5に示す標準モデルDSと人物(例えば顔画像制御装置1のユーザ)の3次元計測データとの概略の位置合わせを行う(#101)。標準モデルDSは、標準的な顔のサイズおよび形状を有した、頭部の全周を構造化した3次元データである。3次元計測データは、点群からなるユーザの顔の3次元データである。すなわち、ステップ#101では、標準モデルDSと3次元計測データとの距離が最小となるように、標準モデルDSの向き、大きさ、および位置を変更する。一般に、標準モデルDSおよび3次元計測データとして、無表情の状態のものが用いられる。なお、3次元計測データは、3次元計測装置でユーザを撮影するなどして予め用意されている。

【0024】輪郭および特徴点を抽出する(#102)。標準モデルDSについての輪郭RKおよび特徴点TTと同じ位置に配置されるべき輪郭および特徴点を、3次元計測データ上に、またはそれに対応する2次元画像上に配置する。

【0025】特徴点として、例えば、目や口の端部、鼻の頂部、顎の下端部のように実際に特徴のある部分、ま

たは、それらの中間のようなそれ自体では特徴はないが位置的に特定し易い部分などが選ばれる。輪郭として、顎のライン、唇のライン、または顔のラインなどが選ばれる。

【0026】計算量および誤差を削減するために、3次元計測データについてデータの削減を行う(#103)。標準モデルDSの変形を行う(#104)。すなわち、3次元計測データの各点と標準モデルDSの面との間の距離に関連して定義されたエネルギー関数、または過剰な変形を回避するために定義されたエネルギー関数などを用い、それらが最小となるように標準モデルDSの面を変形させる。

【0027】そして、対象とするエネルギー関数および制御点を変更し、ステップ#104と同様な変更のための処理を繰り返す(#105)。次に、ステップ#104の変形処理について説明する。

【0028】図7において、3次元計測データを構成する点群の1つが点Pkで示されている。標準モデルDSの面Sにおいて、点Pkに最も近い点がQkで示されている。点Qkは、点Pkから面Sに垂線を下ろしたときの交点である。

【0029】点群に面Sをフィッティングする方法は次の通りである。ここでは、一般的なフィッティングについて説明する。点群の中の1つの点Pk、それに対応する点Qk、および対応点群 $T = \{(Pk, Qk), k = 1 \dots n\}$ について、フィッティングエネルギー(Fitting Energy)関数Ff(U)を、次の式(1)のように設定する。

【0030】

【数1】

$$F_f(U) = \sum_i |P_i - Q_i(U)|^2 \quad \dots (1)$$

【0036】ここで、

【0037】

【数3】

$$\tilde{u}_1^m, \tilde{u}_2^m$$

$$u_1^m, u_2^m$$

【0038】は、それぞれ、仮想バネKBの初期端点、変形後の仮想バネKBの端点である。cはバネ係数であり、Mは仮想バネKBの本数である。また、次の関係が成り立つ。

【0039】

$$F(U) = W_f F_f(U) + W_s F_s(U) \quad \dots (3)$$

【0031】ただし、Qk(U)は、QkがUの関数であることを示す。また、面Sの過度の変形を防ぐために、図8に示す仮想バネ(elastic bar)KBを導入する。仮想バネKBの制約に基づいて、面Sの形状安定化のための安定化エネルギー関数を導く。

【0032】すなわち、図8において、フィッティング対象である標準モデルDSの面(曲面)Sの一部が示されている。面Sは、制御点群 $U = \{u_i, i = 1 \dots n\}$ で形成されている。隣接する制御点間には、仮想バネKBが配置されている。仮想バネKBは、制御点間に引っ張り力による拘束を与え、面Sの異常変形を防ぐ働きをする。

【0033】つまり、隣接する制御点uの間隔が大きくなった場合に、それに応じて仮想バネKBによる引っ張り力が大きくなる。例えば、点Qkが点Pkに近づく場合に、その移動にともなって制御点uの間隔が大きくなると、仮想バネKBによる引っ張り力が增大する。点Qkが移動しても制御点uの間隔が変わらなければ、つまり制御点u間の相対位置関係に変化がなければ、仮想バネKBによる引っ張り力は変化しない。仮想バネKBによる引っ張り力を面Sの全体について平均化したものを、安定化エネルギーとして定義する。したがって、面Sの一部が突出して変形した場合に安定化エネルギーは増大する。面Sの全体が平均して移動すれば安定化エネルギーは零である。

【0034】安定化エネルギー関数Fs(U)は、次の式(2)で示される。

【0035】

【数2】

【数4】

$$L_0^m = \|\tilde{u}_1^m - \tilde{u}_2^m\|$$

【0040】したがって、バネ係数cを大きくすると、仮想バネKBは硬くなって変形し難くなる。このような安定化エネルギー関数Fs(U)を導入することにより、面Sの形状変化に一定の拘束を設けることとなり、面Sの過度の変形を防ぐことができる。

【0041】上に述べたフィッティングエネルギー関数Ff(U)、および安定化エネルギー関数Fs(U)を用い、フィッティングの評価関数F(U)を次の式(3)のように定義する。

【0042】

ここで、 W_f 、 W_s は、それぞれ正規化のための重み係数である。式(3)の評価関数 $F(U)$ が十分小さくなるように、面 S の変形および対応点の探索を繰り返し、面のフィッティングを行う。例えば、 $F(U)$ の U に関する微分が0に近づく方向にフィッティングを行う。

【0043】図6において、変形処理では、まず、点 P_k に対応する点 Q_k を計算で求め、点 P_k と点 Q_k の組みを作成する(#111)。面 S を変形し(#112)、変形後の評価関数 $F(U)$ を計算する(#113)。評価関数 $F(U)$ が収束するまで(#114でYes)、処理を繰り返す。

【0044】評価関数 $F(U)$ の収束を判定する方法として、評価関数 $F(U)$ が所定の値よりも小さくなったときを収束とする方法、前回の計算と比較べた変化の割合が所定値以下となったときに収束とする方法など、公知の方法を用いることが可能である。

【0045】このような処理によって標準モデル DS を変形し、ユーザの顔の形状をした3次元形状モデルを生成することができる。

【筋肉の定義】図9は顔モデル3Mの構成の例を示す図、図10は筋肉配置データ71の例を示す図、図11はノード影響データ72の例を示す図、図12はあるノード N の移動による影響が及ぶ範囲の例を説明する図である。

【0046】顔モデルデータ生成部101によって生成されたユーザの顔の3次元形状モデルは、図3に示す顔

$$N_i = \sum_j V_j' / n$$

N_i : 筋肉 Node i

V_j' : 筋肉 Node i を定義する顔モデルの構成頂点

n : 筋肉 Node i を定義する顔モデルの構成頂点数

【0051】筋肉配置データ71は、図10に示すように、各筋肉(エッジ E_1 、 E_2 、…)の構成に関するデータである。エッジ E の第一のパラメータは、そのエッジ E の端点となる2つのノード N を示す。エッジ E の第二のパラメータは、そのエッジ E (筋肉)を変位させた場合に、どちらの端点(ノード N)をどれだけの割合(ウェイト)で移動させるかを示す。例えば、エッジ E

モデルデータ記憶部102に記憶される。以下、3次元形状モデルを「顔モデル3M」と記載する。また、顔モデルデータ記憶部102は、図10に示す筋肉配置データ71および図11に示すノード影響データ72を記憶する。

【0047】図9(a)において、複数の細い直線同士の交点は、顔モデル3Mの構成頂点(Model Vertex) V を示す。顔の表面すなわち皮膚の位置は、構成頂点 V によって決まる。

【0048】太い直線は、顔モデル3Mの筋肉を意味するエッジ(Edge) E を示す。黒い丸印は筋肉の端点を意味するノード N (Node)を示す。つまり、筋肉(エッジ E)の位置は、異なる2つのノード N によって決まる。ノード N (N_1 、 N_2 、…)は、3次元形状モデルの各制御点 u に対応しており、顔全体の各筋肉の端点となる位置に配置されている。なお、図9(b)は、ノード N とエッジ E との関係を分かりやすくするために図9

(a)から構成頂点 V を省略して示している。図9

(a)(b)は、顔の右半分のノード N およびエッジ E を省略して示しているが、実際には、左半分と同様にノード N およびエッジ E が存在する。

【0049】ノード N の位置は、次に示す式(4)のように構成頂点 V の相対的位置として表される。

【0050】

【数5】

$$\dots (4)$$

3の第二のパラメータ「0.7、0.3」は、ノード N_4 を7、ノード N_3 を3、の割合でそれぞれ移動させることを示している。

【0052】エッジ E が変位するとき、ノード N が移動する位置は、次に示す式(5)によって求められる。

【0053】

【数6】

$$N_1^{t'} = w_1^t t^t \overrightarrow{N_1^t N_2^t} + N_1^t \quad \dots (5)$$

N_1^t : 変位前の Edgei の第 1 端点 Node

N_2^t : 変位前の Edgei の第 2 端点 Node

$N_1^{t'}$: 変位後の Edgei の第 1 端点 Node

w_1^t : Edgei の第 1 端点 Node のウェイト

t^t : Edgei の変位量

【0054】ただし、実際には複数のエッジEに関係するノードNが存在するため、収束演算または連立演算によってノードNの移動後の位置が求められる。ノード影響データ72は、図11に示すように、ノードNの移動に伴って構成頂点Vに及ぼされる影響に関するデータである。ノード影響データ72の第二のパラメータは、各ノードNが移動したときに影響を受ける構成頂点Vを示している。つまり、ノードNが移動したときの影響の範囲を示している。ノードNの移動による影響を受ける構成頂点Vは、そのノードNの周辺に集中している。例えば、図12において、大きい黒丸が示すノードNの移動による影響を受ける構成頂点Vは、小さい黒丸が示す9つの構成頂点Vである。

【0055】第一のパラメータは、ノードNが移動したときに構成頂点Vに対して与える影響の度合 (intensity) を示している。この値が大きいと、ノードNの移動に伴う構成頂点Vの移動量 (変位量) が大きくなる。

【0056】ノードNが移動するのに伴って構成頂点Vが移動する位置は、次に示す式 (6) によって求められる。

【0057】

【数7】

$$V_i = \sum_j w_j^t N_j^t / d_j^t \quad \dots (6)$$

V_i : 移動する顔モデルの構成頂点

N_j^t : V_i に影響する筋肉 Node

w_j^t : 筋肉 N_j^t が V_i に与える影響度合

d_j^t : V_i と N_j^t の距離

【0058】〔顔モデルの形状を制御するためのデータ〕図13は母音および子音が通常属する形状グループを示す図、図14は形状グループデータ75の例を示す図、図15は形状変化データ74の例を示す図、図16は各形状グループに属する音を発したときの顔モデルの形状の例を示す図である。

【0059】本実施形態では、図13に示すように、顔モデル3Mの形状の変化に関するデータ量の軽減のため

に、5つの母音のグループ (形状グループA、E、I、O、U) および発音する際の口の形状の特徴に基づいて各子音の種類を分類するための3つのグループ (形状グループ1~3) が設けられている。

【0060】形状グループA、E、I、O、Uには、それぞれ「a」、「e」、「i」、「o」、「u」の1種類ずつの母音が属する。形状グループ1は唇を合わせて発音する子音のグループ、形状グループ2は唇を合わせずに口を所定の形状にして発音する子音のグループ、形状グループ3は前に発した音の口の形状のまま発音する子音のグループである。係る分類によると、通常、形状グループ1には「b、f、m、p、v」の5種類の子音が属し、形状グループ2には「d、g、j、k、l、n、r、s、t、w、z」の11種類の子音が属し、形状グループ3には「h、y」の2種類の子音が属する。

【0061】図3の形状変化データ記憶部107は図15に示す形状変化データ74を記憶する。形状変化データ74は、各形状グループA、E、I、O、U、1、2に属する音を発するとき、顔モデル3Mがそれぞれ図16(a)~(g)に示す形状になるように顔モデル3Mの各筋肉 (エッジE1、E2、...) を変化させるためのデータである。形状変化データ74の各値は、筋肉の収縮の度合を示している。収縮していない状態を

「0」、最も収縮した状態を「20」とする。例えば、変形量 (収縮の度合) が「15、0」であれば、その筋肉 (エッジE) が75%収縮することを示す。なお、形状グループ3の場合は、顔モデル3Mは前に発した音 (音韻) の形状のまま保たれるので、変形量の値を持たない。

【0062】後に説明する顔モデル3Mの形状の制御は、その顔モデル3Mに喋らせる言葉を音韻ごとに区切って分解し、各音韻が属する形状グループに対応する形状変化データ74に基づいてエッジEを変化させることによって行われる。例えば、「館 (kan)」という言葉の場合は、「k、a、n」のように区切られ、順に形状グループ2、A、2の形状になるように顔モデル3Mを制御する。

【0063】しかし、一般に、言葉には、直前の音 (音韻) の影響を受けて記述 (スペル) 通りに発音しない場

合がある。例えば、「公(k o u)」という言葉の場合は、「u」がその直前の「o」の影響を受けて「k o o」と発音される。このとき、口の形状も、図16(e)に示す「u」の形状ではなく、図16(d)に示す「o」の形状となる。

【0064】形状グループ記憶部108に記憶される形状グループデータ75は、図14に示すように、上記の例のように直前の音韻の影響を受けて通常とは異なる口の形状で発音される音韻に関するデータである。具体的には、形状グループデータ75は、各音韻がその直前の音韻の影響を受けるとどの形状グループに変更されるかを示している。例えば、「g」は通常は形状グループ2に属するが、直前に「n」があるときは太字で示すように形状グループ3に変更される。

〔顔画像の制御(アニメーションの実行)〕次に、ユーザの顔画像HFの動作の制御について説明する。

【0065】図17は「k o u m i n k a n」の各音韻が属する形状グループの例を示す図、図18は母音、子音、および連続母音の継続時間を説明する図、図19はタイムテーブルの例を示す図である。

【0066】図3のテキストデータ取得部103は、顔モデル3Mの形状を変形する基となる言葉(すなわち顔モデル3Mに喋らせる言葉)をテキストデータ73として取得する。例えば、ユーザがキーボード13を操作して入力した文字列をテキストデータ73として取得する。または、他のユーザから受け取った電子メールに書かれているメッセージをテキストデータ73としてもよい。

【0067】テキストデータの代わりに、音声によって言葉を入力してもよい。音声データ取得部104は、マイク15に向かって喋ったユーザの声を音声データ74として取得する。音声データ74は、音声テキスト変換部105によってテキストデータ73に変換される。

【0068】顔画像制御部106は、形状取得部161、時間配分部162、形状補間部163、および動画画像生成部164によって構成され、入力された言葉に合わせて顔画像HFが動作するように制御する。

【0069】形状取得部161は、テキストデータ73を音韻すなわち母音または子音ごとに区切って分解する。例えば、テキストデータ73が「公民館(こうみんかん)」である場合は、「k、o、u、m、i、n、k、a、n」の9つの音韻に区切られる。そして、区切られた各音韻がいずれの形状グループに属するかを、その直前の音韻および図14に示す形状グループデータ75に基づいて求める。すなわち、各音韻について、直前の音韻の影響を受けて形状グループの変更があるか否かを

形状グループデータ75に基づいて求める。例えば、図17に示すように、3番目の「u」は、直前に「o」があるので形状グループOに変更される。これ以外の音韻は、直前の音韻の影響を受けないので、通常の形状グループのままである。このようにして、図17に示すように、各音韻「k、o、…、n」に合わせて顔モデル3Mの形状を制御するための形状グループが求められる。ただし、図17において、「G1」および「G2」はそれぞれ形状グループ1および2を示し、「A」、「I」、および「O」はそれぞれ形状グループA、I、およびOを示している。

【0070】時間配分部162は、各音韻に合わせて顔モデル3Mの形状を変形させる際のタイミングに関する設定を行う。係る設定は、次に示す規則に基づいて行われる。

【0071】図18に示すように、ある形状(例えば無表情の形状)から次の音韻を発した形状に変化するまでの時間(立ち上がり時間T1)およびその形状が無表情の形状に戻るまでの時間(終息時間T3)を極めて短い時間とし、例えば、0.1秒以下とする。このとき、 $T1 = T3$ としてもよい。

【0072】母音を発した形状を保つ時間(継続時間Tb2)、子音を発した形状を保つ時間(継続時間Ts2)、および直前の音韻が母音である場合の母音(連続母音)を発した形状を保つ時間(継続時間Tr2)の長さの関係を、 $Ts2 < Tr2 < Tb2$ 、とする。標準的な会話のスピードであれば、Ts2がおおよそ0.1秒、Tb2がおおよそ0.4秒くらいとなる。これらの時間の長さは、話すスピードに応じて変更可能である。

【0073】隣り合う2つの音韻のタイミングは、図19に示すように、前の音韻が終息したとき($t = tb$)に後の音韻の立ち上がりが完了するように設定する。つまり、後の音韻は、tbよりも終息時間T3だけ前に立ち上がりはじめるように配置される($t = ta$)。

【0074】上記の規則によると、例えば、「k o o m i n k a n」の各音韻に対応する形状グループが、図19に示すタイムテーブルのように配置される。図3に戻って、形状補間部163は、ある音韻からその次の音韻へ移る際の顔モデル3Mの形状のデータを補間する。つまり、一方の音韻の立ち上がり時間と他方の終息時間とが重なる部分の顔モデル3Mの形状を簡単のため直線近似とし、次の式(7)に基づいて顔モデル3Mの各構成頂点Vの位置を求める。

【0075】

【数8】

$$V_i(t) = c_a V_i(ta) + c_b V_i(tb) \quad \dots (7)$$

$V_i(t)$: 時刻 t の顔モデルの構成頂点 V_i の位置

$$c_a = (t - ta) / (tb - ta)$$

$$c_b = (tb - t) / (tb - ta)$$

【0076】動画像生成部164は、形状取得部161および形状補間部163によって得られた顔モデル3Mの形状すなわち構成頂点Vの位置を図19に示すタイムテーブルに従って変化させながら所定の方向から2次元上に投影することによって顔画像HFを動作させ、リップシンクアニメーションを生成する。

【0077】音声出力部109は、言葉を音声化し、リップシンクアニメーションと同期して出力する。例えば、所定の音韻が立ち上がるときに顔画像制御部106から発せられる信号(トリガー)に合わせて順次音声を出力する。テキストデータを音声化する方法として、公知の音声合成技術が用いられる。

【0078】次に、入力された言葉に合わせてユーザの顔画像HFを動かしてリップシンクアニメーションを生成する処理の流れを、フローチャートを参照して説明する。図20は顔画像制御装置1の全体の処理の流れを説明するフローチャート、図21は形状の取得の処理の流れを説明するフローチャート、図22は時間配置の処理の流れを説明するフローチャートである。

【0079】図20に示すように、まず、ユーザの顔画像の基となる顔モデル3Mを生成し(#1)、エッジE(筋肉)の位置を設定する(#2)。ノードNの位置の変化による影響が及ぶ構成頂点Vの範囲を設定し、顔モデル3Mの筋肉が動いたときの皮膚への影響について設定する(#3)。形状グループごとにエッジEの位置を設定する(#4)。ステップ#1~#4は、顔画像の生成のための準備なので、既に顔モデル3Mを生成し各設定を行っている場合は省略してもよい。

【0080】ユーザの顔画像に喋らせる言葉をテキストデータとして入力する(#5)。音声で入力した場合は、テキストに変換する。入力したテキストデータに含まれる音韻ごとに顔モデル3Mの形状を取得する(#6)。すなわち、図21に示すように、テキストデータを音韻ごとに区切って分解し(#121)、図14に示す形状グループデータ75に基づいて各音韻が属する形状グループを求め(#122)、求められた各形状グループに対応する顔モデル3Mの形状を図15に示す形状変化データ74に基づいて取得する(#123)。

【0081】分解された各音韻をタイムテーブルに配置する(#7)。すなわち、図22に示すように、音韻が子音である場合は(#131でNo)その音韻の継続時間をTs2とし(#132)、連続母音である場合は(#133でYes)継続時間をTr2とし(#134)、通常の母音である場合は(#133でNo)継続

時間をTb2とする(#135)。

【0082】タイムテーブルに配置された隣り合う音韻同士の重なる部分について、形状の補間を行う(#8)。そして、上記の処理によって得られたデータに基づいて顔モデル3Mの形状を変形させ、リップシンクアニメーションを生成する(#9)。

【0083】本実施形態によると、言葉の音を音韻ごとすなわち母音または子音ごとに区切り、区切られた各音韻ごとに顔画像の動作を制御するので、従来よりもデータの量を削減しつつリアルなリップシンクアニメーションを実現することができる。さらに、本実施形態では、5つの母音の形状グループおよび3つの子音の形状グループを設け、子音の種類を3つの子音の形状グループのうちのいずれかに分類している。したがって、形状に関するデータを8種類だけ準備すればよいので、一層データの量を削減することができる。

【0084】また、本実施形態では、子音の継続時間Ts2を母音の継続時間Tb2よりも短く設定し、顔画像の形状の変化を制御している。これにより、よりリアルなリップシンクアニメーションを実現することができる。また、隣り合う2つの音韻の間について形状のデータの補間処理を行っているので、音韻間の形状の変化が統合され、全体として滑らかに変化するリップシンクアニメーションを実現することができる。

【0085】本実施形態では、ユーザの3次元計測データに標準モデルをフィッティングすることにより顔モデルを取得したが、ユーザの2次元画像に標準モデルをフィッティングして顔モデルを取得してもよい。または、種々のコンピュータグラフィック(CG)プログラムを用いて顔モデルを作成してもよい。

【0086】言葉に合わせて顔画像の口を動かすだけでなく、目、眉、または表情などを変化させるようにしてもよい。例えば、時間ごとの目、眉、または表情などの変化を表すデータを用意し、係るデータに基づいて顔モデルの筋肉(エッジE)の制御を行うようにする。

【0087】本実施形態の顔画像制御装置1は、複数のユーザ間におけるコミュニケーション手段として利用可能である。例えば、ユーザαとユーザβとが画像処理機能を備えた携帯電話で会話をする場合に、ユーザα、βは自分の顔モデルを相手の携帯電話装置に予め送っておく。ユーザβの携帯電話端末は、受信したユーザαの音声を音韻ごとに区切り、図20に示すステップ#6~#9の処理を行い、ユーザαのリップシンクアニメーションを表示する。同様に、ユーザαの携帯電話端末は、ユ

ーザ β のリップシンクアニメーションを表示する。

【0088】本実施形態では、3次元の顔モデルを用いて顔画像の制御を行ったが、2次元の顔モデルであってもよい。本実施形態では、子音の種類を3つの形状グループに分類したが、4つ以上の形状グループに分類してもよい。また、英語などの場合は、テキストデータを発音記号に変換し、発音記号に基づいて顔画像を制御するようにしてもよい。

【0089】その他、顔画像制御装置1の全体または各部の構成、処理内容、処理順序などは、本発明の趣旨に沿って適宜変更することができる。

【0090】

【発明の効果】本発明によると、従来よりも形状に関するデータの量を少なくし、かつ、顔画像の制御によるリップシンクアニメーションをリアルに実現することができる。

【図面の簡単な説明】

【図1】本発明に係る顔画像制御装置の構成を説明する図である。

【図2】磁気記憶装置に記憶されるプログラムおよびデータを示す図である。

【図3】顔画像制御装置の機能的構成を説明する図である。

【図4】3次元形状モデルの生成の処理の流れを説明するフローチャートである。

【図5】標準モデルの例を示す図である。

【図6】変形処理の流れを説明するフローチャートである。

【図7】標準モデルの面Sと3次元計測データの点Pとを模式的に示す図である。

【図8】標準モデルの異常変形を防ぐための仮想パネを

説明するための図である。

【図9】顔モデルの構成の例を示す図である。

【図10】筋肉配置データの例を示す図である。

【図11】ノード影響データの例を示す図である。

【図12】あるノードの移動による影響が及ぶ範囲の例を説明する図である。

【図13】母音および子音が通常属する形状グループを示す図である。

【図14】形状グループデータの例を示す図である。

【図15】形状変化データの例を示す図である。

【図16】各形状グループに属する音を発したときの顔モデルの形状の例を示す図である。

【図17】「kouminkan」の各音韻が属する形状グループの例を示す図である。

【図18】母音、子音、および連続母音の継続時間を説明する図である。

【図19】タイムテーブルの例を示す図である。

【図20】顔画像制御装置の全体の処理の流れを説明するフローチャートである。

【図21】形状の取得の処理の流れを説明するフローチャートである。

【図22】時間配置の処理の流れを説明するフローチャートである。

【符号の説明】

1 顔画像制御装置

106 顔画像制御部（動作制御手段）

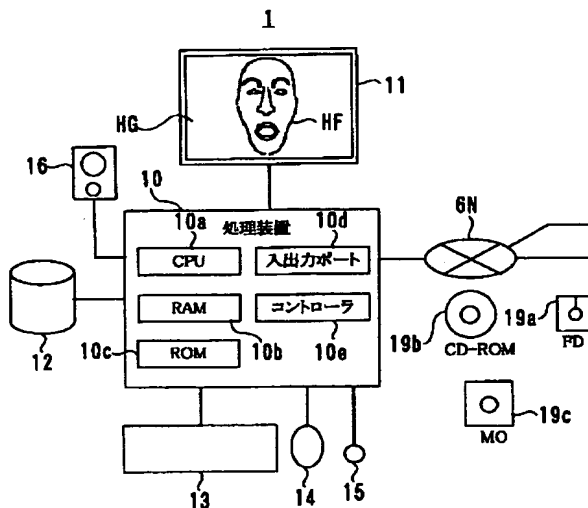
107 形状変化データ記憶部（形状データ記憶手段）

19a～19c リムーバブルディスク（記録媒体）

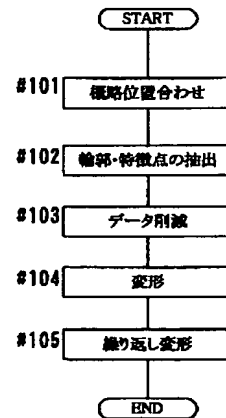
74 形状変化データ（形状データ）

HF 顔画像

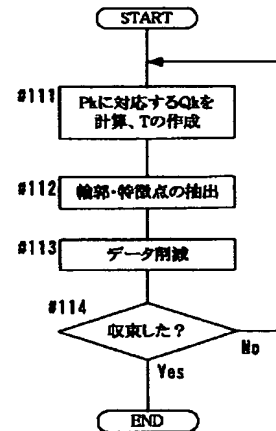
【図1】



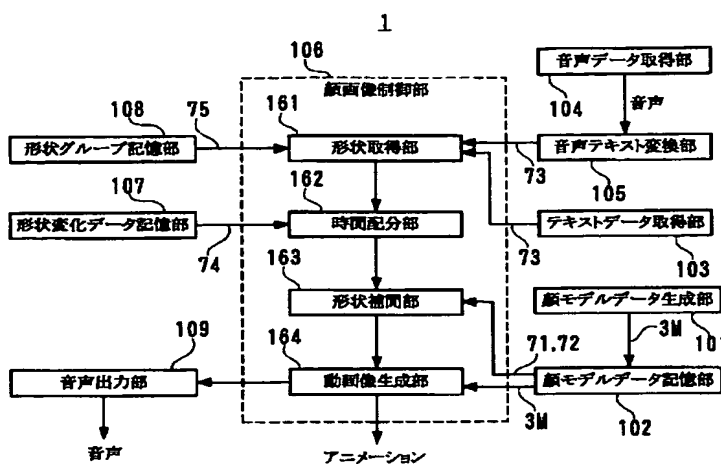
【図4】



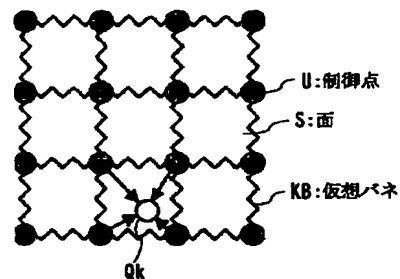
【図6】



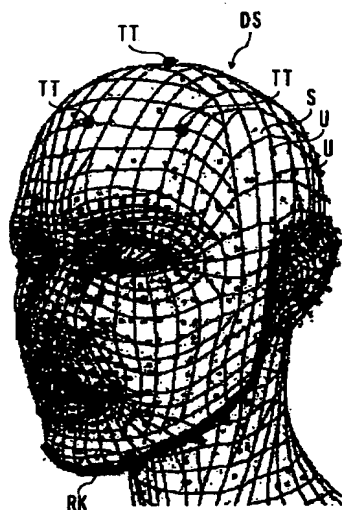
【図3】



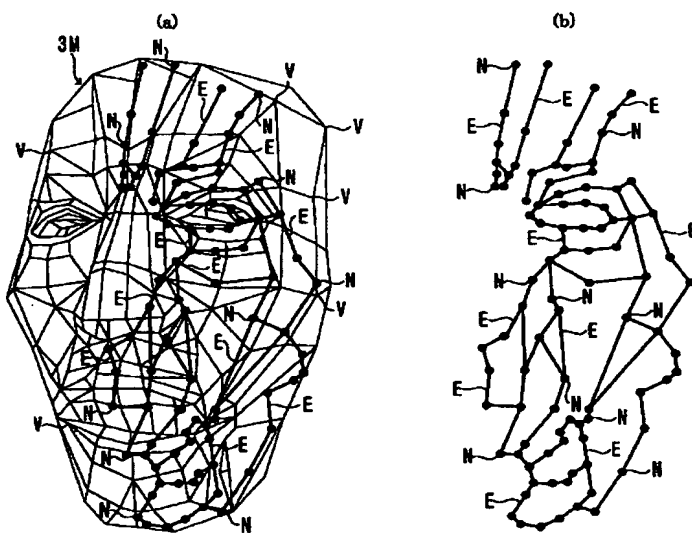
【图8】



【図5】



【図9】



【図10】

71			
TYPE	NAME	第一のパラメータ	第二のパラメータ
Node	N1	50,43,15	
Node	N2	33,25,29,41	
Node	N3	1,5,27	
Edge	E1	N1,N2	1.0,0.0
Edge	E2	N2,N3	1.0,0.0
Edge	E3	N4,N5	0.7,0.3

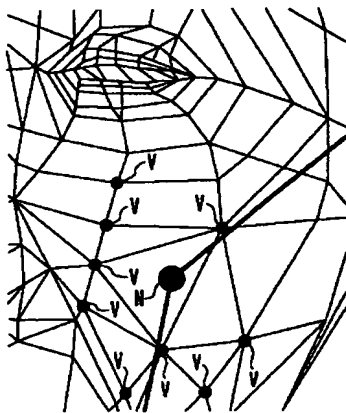
【図11】

72		
Node	Intensity	Model Vertex
N1	10.0	50,43,15,44,45,46
N2	2.5	33,25,29,41,26,27,28,30,31,32
N3	5.0	1,5,27,2,3,4

【図13】

Group	音韻(母音、子音)
A	a
E	e
I	i
O	o
U	u
1	b,f,m,p,v
2	d,g,j,k,l,n,r,s,t,w,z
3	h,y

【図12】



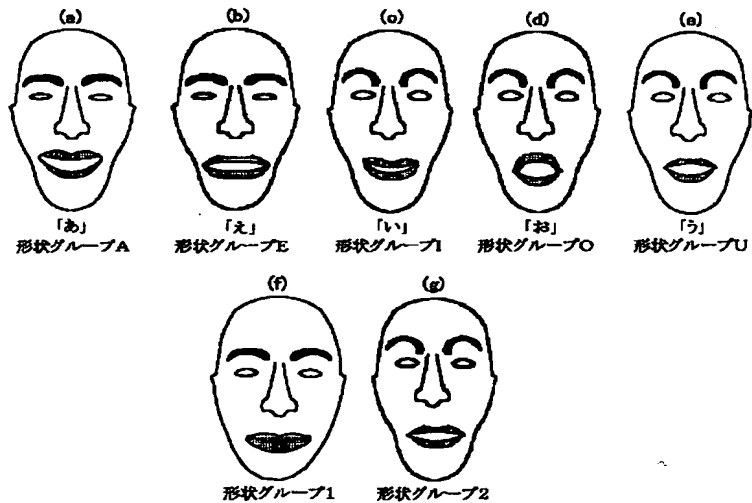
【図14】

75				
	音韻			
	通常	「o」の後	「n」の後	「a,e,i,o,n,y」の後
グループA	a	a	a	a
グループB	e	e	e	e
グループI	i	i	i	i
グループO	o	o,u	o	o
グループU	u	-	u	u
グループ1	b,f,m,p,v	b,f,m,p,v	b,f,m,p,v	b,f,m,p,v
グループ2	d,g,j,k,l,n,r,s,t,w,z	d,g,j,k,l,n,r,s,t,w,z	d,g,j,k,l,n,r,s,t,w,z	d,g,j,k,l,n,r,s,t,w,z
グループ3	h,y	h,y	h,g,y	h,y

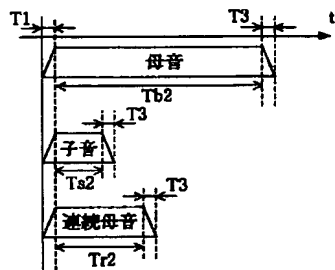
【図15】

	Edge ID			
	E1	E2	E3	...
グループA	0.0	0.0	15.0	...
グループE	3.0	0.0	4.0	...
グループI	6.0	6.0	0.0	...
グループO	3.6	2.2	15.0	...
グループU	2.2	1.7	1.0	...
グループ1	0.0	0.0	0.0	...
グループ2	0.0	0.0	6.0	...
グループ3	-	-	-	...

【図16】



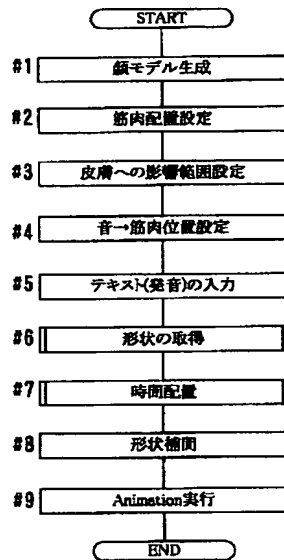
【図18】



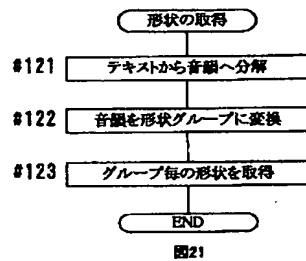
【図17】

	1	2	3	4	5	6	7	8	9
音韻	k	o	u	m	i	n	k	a	n
直前の音韻	-	k	o	u	m	i	n	k	a
変更	-	-	o	-	-	-	-	-	-
グループ	G2	O	O	G1	I	G1	G2	A	G1

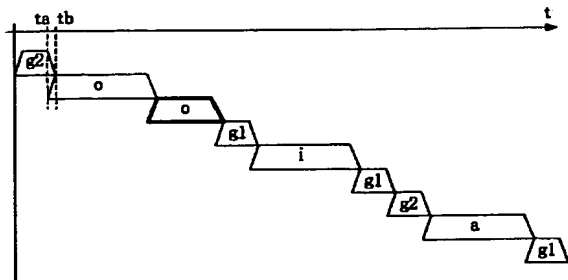
【図20】



【図21】



【図19】



【図22】

